Title

Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site

Authors

Javier A. Jaimes[1], Jean K. Millet[2], Gary R. Whittaker[1,3*]

Affiliations

[1]Department of Microbiology & Immunology, Cornell University, Ithaca NY 14853, United States;

[2]Université Paris-Saclay, INRAE, UVSQ, Virologie et Immunologie Moléculaires, 78352 Jouy-en-Josas, France;

[3]Master of Public Health Program, Cornell University, Ithaca NY 14853, United States.

*- contact author

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 19 (COVID-19) has rapidly spread from an initial outbreak in Wuhan, China in December 2019 to the rest of the world within a few months. On March 11th 2020, the rapidly evolving COVID-19 situation was characterized as a pandemic by the WHO. Much attention has been drawn to the origin of SARS-CoV-2, a virus which is related to the lineage B betacoronavirus SARS-CoV and SARS-related coronaviruses found in bat species. The closest known relative to SARS-CoV-2 is a bat coronavirus named RaTG13 (BatCoV-RaTG13). Early characterizations of the SARS-CoV-2 genome revealed the existence of a distinct 4 amino acid insert (underlined, SPRRAR↓S), found within the spike (S) protein, at a position termed the S1/S2 site located at the interface between the S1 receptor binding subunit and the S2 fusion subunit. Notably, this S1/S2 insert appears to be distinguishing feature among SARS-related sequences and introduces a potential cleavage site for the protease furin. Here, we investigate the potential role of this novel S1/S2 cleavage site and present direct biochemical evidence for proteolytic processing by a variety of proteases, including furin, trypsin-like proteases and cathepsins. We discuss these findings in the broader context of the origin of SARS-CoV-2, viral stability and transmission.

Introduction

Since December 2019, human infections by a novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have rapidly spread globally from an initial outbreak of severe pneumonia centered on Wuhan, Hubei Province, China. The virus is the etiological agent of an infectious respiratory disease termed coronavirus disease 19 (COVID-19). On March 11th 2020, due to its alarming spread and severity across most countries globally the WHO characterized COVID-19 as a pandemic.

Initial genomic characterization and phylogenetic analyses have revealed that SARS-CoV-2 is related to the lineage B betacoronavirus SARS-CoV, which caused the 2002-2003 SARS epidemic. Bats are considered as the likely source of SARS-CoV and SARS-CoV-2 viruses as a rich diversity of SARS-related (or SARS-like) viruses have been identified in several bat species. Notably, a bat coronavirus named BatCoV-RaTG13 has been identified as being most closely related to SARS-CoV-2 on a genomic level with a remarkable 96% sequence identity (Zhou et al., 2020b). To date, the direct precursor of SARS-CoV-2 and the involvement of an intermediate host in the emergence of the human virus remain to be identified and there has been much attention drawn to the origin of the virus.

One of the notable features of the SARS-CoV-2 genome is a four amino acid insert (shown underlined below) found at the S1/S2 junction of the SARS-CoV-2 spike protein. Following the release of the SARS-CoV-2 genome sequence, several groups including ourselves identified this insert (underlined, S<u>PRRA</u>R↓S) as a potential cleavage site for the protease furin (Coutard et al., 2020; Jaimes et al., 2020; Walls et al., 2020; Wrapp et al., 2020), which is widely used to activate the fusion machinery of viral glycoproteins. This insert has also been referred to as a “polybasic site” (Andersen et al., 2020). However, trypsin-like proteases (notably TMPRSS2) and cathepsins have recently been shown to be functionally important for SARS-CoV-2 infection (Hoffmann et al., 2020; Ou et al., 2020). Previous studies on SARS-CoV, MERS-CoV, and other coronaviruses have shown that activation of the spike protein is often a complex process involving multiple cleavage events occurring at distinct sites and with the involvement of several host proteases (Hulswit et al., 2016). Moreover, work on SARS-CoV has also demonstrated the criticality of a cleavage event occurring at another site termed S2’ (Belouzard et al., 2009),

found in the S2 fusion domain immediately upstream of the fusion peptide—whose functional role is more directly comparable to that of influenza virus HA cleavage site (Steinhauer, 1999).

Furin is ubiquitously expressed in the Golgi apparatus of all cells, but generally only at low levels—with some cell types showing enhanced expression or altered intracellular distribution (Shapiro et al., 1997). As seen with other coronaviruses (Le Coupanec et al., 2015; Licitra et al., 2013; Millet and Whittaker, 2014), the role of furin (or other proteases) is potentially highly significant, but needs to be carefully interpreted in the context of other aspects of virus infection. The role of furin cleavage as a virulence factor for viruses is rooted in findings on influenza virus, where a "polybasic" or "multi-basic" site was found to be a defining feature of highly pathogenic avian influenza (HPAI) (Kawaoka and Webster, 1988), as adopted by the World Organization for Animal Health (OIE). In a subset of avian influenza virus subtypes (*i.e.* H7 and H5) polymerase slippage inserts a stretch of purine residues at the interface of the HA1 and HA2 subdomains, based on a specific secondary RNA structure found in the H7 and H5 HA gene (Nao et al., 2017). Such polybasic sites typically exist as a stretch of 6-7 arginine (R) and lysine (K) residues (*e.g.* RKKRKR↓G) that can be cleaved by furin, therefore allowing systemic spread based on the ubiquitous expression of the protease. Without the polybasic cleavage site, infection is restricted based on the localized presence of the trypsin-like protease activating low pathogenicity influenza viruses. However, other influenza viruses (notably H9) can modulate cleavage sites in a distinct and less well-recognized way, creating a more defined (but not necessarily "polybasic") furin cleavage site through recombination and mutation (Lee and Whittaker, 2017).

Here we provide context and clarify the role of the novel SARS-CoV-2 S1/S2 cleavage site in virus emergence and infection, and perform a direct assessment of the proteases cleaving this site by use of biochemical assays.

Results

To directly address the proteases cleaving the SARS-CoV-2 S1/S2 site, we used a biochemical peptide cleavage assay (Jaimes et al., 2019), which was previously used to screen emerging influenza viruses (Straus and Whittaker, 2017). This assay was also successfully used to study the functional impact of mutations at

coronavirus spike cleavage sites for feline coronavirus and MERS-CoV (Licitra et al., 2013; Millet et al., 2016). The peptide sequences used here were **HTVSLLRSTSQ** (SARS-CoV S1/S2) and **TNSPRRARSVA** (SARS-CoV-2 S1/S2). We tested a range of proteases likely to be involved in spike protein processing–the proprotein convertases furin and PC1, trypsin and the type II transmembrane serine protease (TTSP) matriptase, as well as cathepsins B and L. As predicted, furin cleaves SARS-CoV-2, but not SARS-CoV (fig. 1). However, in addition to furin, other proteases also cleaved SARS-CoV-2 much more readily than SARS-CoV. PC1 showed a similar cleavage pattern but with lower efficiency than for furin. Trypsin cleaved both peptides, but was over 4-fold more efficient for SARS-CoV-2. Both the TTSP matriptase and cathepsin B did not cleave SARS-CoV at all, but were highly active on SARS-CoV-2. The only protease that showed more cleavage on SARS-CoV compared to SARS-CoV-2 was cathepsin L. Our data demonstrate that the S1/S2 site of SARS-CoV-2 S is efficiently cleaved by a wide range of proteases, not only furin. The comparative data with SARS-CoV S1/S2 site reveals that the acquisition of the 4 amino acid insert distinctively broadens the activating protease repertoire of the SARS-CoV-2 S1/S2 cleavage site to all major classes of proteolytic enzymes known to potentially activate coronavirus S proteins.

Discussion

The COVID-19 pandemic represents a global public health emergency. The origin of the causative agent of the disease, SARS-CoV-2 remains a mystery. Previous analyses have revealed distinctive features within the genome of SARS-CoV-2. In particular, a 4 amino acid insert was found within the S1/S2 site of the S glycoprotein. This finding has garnered interest because of the possible introduction of a furin recognition motif at the S1/S2 site. Our study aimed at clarifying the functional role of the S1/S2 cleavage site for SARS-CoV-2. We provide direct biochemical evidence that the site S1/S2 is recognized and cleaved by furin. However, other proteases such as PC1 another member of the pro-protein convertase (PC) family of proteases, trypsin-like proteases, and cathepsins can all efficiently recognize and cleave SARS-CoV-2 S1/S2 cleavage site. These data confirm earlier findings showing the SARS-CoV-2 entry was dependent, at least in part, on lysosomal cathepsins and the cell surface expressed type II transmembrane serine protease (TTSP) TMPRSS2 (Hoffmann et al., 2020). Interestingly,

126 the comparative assessment we present with SARS-CoV shows that the S1/S2 insertion SARS-CoV-2 has acquired
127 substantially expands its proteolytic activation profile, with potential activation from a wide variety of proteases.
128 Another feature of the S1/S2 junction that has been noted for SARS-CoV-2 is the presence of a leading proline
129 (P) residue. We favor a structural interpretation for the role of this residue, whereby the presence of the leading
130 proline—and the turn it creates—impacts the cleavability of the S1/S2 junction by directly affecting the
131 accessibility of the cleavage loop with the active site of the protease. It has also been suggested that the leading
132 proline residue allows the addition of O-linked glycans to neighboring residues (Andersen et al., 2020), with the
133 suggestion that these glycans act to positively influence virulence by creating a mucin-like domain that shields
134 critical epitopes from the immune system. While this is possible, such glycan addition would also be predicted
135 to shield the cleavage site from its activating protease, so negatively influencing virulence. Such a scenario (in
136 this case via N-linked glycans) has been noted for influenza HA (Tse et al., 2014). Not all algorithms predict O-
137 glycans in this region and it remains to be determined whether the precited glycosylation sites are in fact utilized.
138 We favor a model whereby the novel S1/S2 insert allows furin cleavage but also reflects a more enhanced
139 exposure of a critical cleavage site to TTSPs such as TMPRSS2 and matriptase, as well as cathepsin B, an activation
140 mechanism that is not necessarily through the action of furin itself. Overall, the novel S1/S2 insert is likely to
141 enhance spike protein cleavage by multiple proteases beyond that for other SARS-like viruses (*e.g.* containing
142 the sequence SR↓S). The role of the fusion peptide-proximal S2' site remains to be evaluated, but its sequence
143 (PDPSKPSKR↓SFIEDLLF) is not distinctly different from other SARS-like viruses (Jaimes et al., 2020). We also
144 propose that the SARS-CoV-2 S1/S2 cleavage site likely arose as for influenza H9 viruses, rather than as for the
145 more widely appreciated polybasic sites found in H7 and H5 HPAI, and that the term "polybasic site" is a
146 misnomer for SARS-CoV-2. This concept has recently been reinforced by the finding of a bat coronavirus (BatCoV-
147 RmYN02) closely related to SARS-CoV-2 that has an extended S1/S2 cleavage loop lacking any additional basic
148 residues (SPAAR↓S) (Zhou et al., 2020a). BatCoV-RmYN02 represents a good candidate to be involved in the
149 emergence of SARS-CoV-2; lacking a furin cleavage site and likely utilizing a distinct receptor, but providing
150 evidence for a natural origin of SARS-CoV-2.

Our work underscores the importance of biochemical validation of newly identified coronavirus spike glycoprotein cleavage sites to gain a better understanding of the potential functional implications of proteolytic activation. Further investigation is required to identify precursor sequences that led to the emergence of SARS-CoV-2 and to shed light on the evolutionary mechanism(s) by which the virus acquired its distinct S1/S2 site.

Methods

Fluorogenic peptides derived from SARS-CoV and SARS-CoV-2 spike (S) S1/S2 sites composed of the sequences HTVSLLRSTSQ and TNSPRRARSVA sequences, respectively, and harboring the (7-methoxycoumarin-4-yl)acetyl/2,4-dinitrophenyl (MCA/DNP) FRET pair were synthesized by Biomatik (Wilmington, DE, USA). Recombinant furin was purchased from New England Biolabs (Ipswich, MA, USA). Recombinant L-1-Tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin was obtained from Sigma-Aldrich (St Louis, MO, USA). Recombinant PC1, matriptase, cathepsin B , and cathepsin L were purchased from R&D Systems (Minneapolis, MN, USA).

For each fluorogenic peptide, a reaction was performed in a 100 µL volume with buffer composed of 100 mM Hepes, 0.5% Triton X-100, 1 mM $CaCl_2$ and 1 mM 2-mercaptoethanol pH 7.5 for furin (diluted to 10 U/mL); 25 mM MES, 5 mM $CaCl_2$, 1% (w/v) Brij-35, pH 6.0 for PC1 (diluted to 2.2 ng/µL); PBS for trypsin (diluted to 8 nM); 50 mM Tris, 50 mM NaCl, 0.01% (v/v) Tween® 20, pH 9.0 for matriptase (diluted to 2.2 ng/µL); 25 mM MES, pH 5.0 for cathepsin B (diluted to 2.2 ng/µL); 50 mM MES, 5 mM DTT, 1 mM EDTA, 0.005% (w/v) Brij-35, pH 6.0 for cathepsin L (diluted to 2.2 ng/µL) and with the peptide diluted to 50 µM. Reactions were performed at 30 °C in triplicates, and fluorescence emission was measured every minute for 45 min using a SpectraMax fluorometer (Molecular Devices, Sunnyvale, CA, USA), with λ_{ex} 330 nm and λ_{em} 390 nm wavelengths setting, enabling tracking of fluorescence intensity over time and calculation of Vmax of reactions. Assays were performed in triplicates with results representing averages of Vmax from three independent experiments.

Figure legend

Figure 1. Proteolytic cleavage assay of peptides derived from SARS-CoV and SARS-CoV-2 S1/S2 spike protein sites. Fluorogenic peptide mimetics of the S1/S2 spike cleavage sites of SARS-CoV (HTVSLLRSTSQ) and SARS-CoV-2 (TNSPRRARSVA) were incubated with the following recombinant proteases (indicated on top of graphs): furin, PC1, trypsin, matriptase, cathepsin B, and cathepsin L. The increase in fluorescence due to proteolytic processing was measured using a fluorometer enabling calculation of the velocity of each cleavage reactions (Vmax). The assays were performed in triplicates with results representing Vmax averages obtained from three independent experiments (n=3). Error bars indicate SD.

Acknowledgements: We thank Ruth Collins and all member of the Whittaker and Daniel labs at Cornell University for comments and discussion. Work in the author's laboratory is supported by the National Institutes of Health (research grant R01AI35270).

References

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine.

Belouzard, S., Chu, V.C., and Whittaker, G.R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proceedings of the National Academy of Sciences *106*, 5871–5876.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res *176*, 104742.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A*., et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.

Hulswit, R.J., de Haan, C.A., and Bosch, B.J. (2016). Coronavirus Spike Protein and Tropism Changes. Adv Virus Res *96*, 29-57.

Jaimes, J.A., André, N.M., Millet, J.K., and Whittaker, G.R. (2020). Structural modeling of 2019-novel coronavirus (nCoV) spike protein reveals a proteolytically-sensitive activation loop as a distinguishing feature compared to SARS-CoV and related SARS-like coronaviruses. bioRxiv, 2020.2002.2010.942185.

Jaimes, J.A., Millet, J.K., Goldstein, M.E., Whittaker, G.R., and Straus, M.R. (2019). A Fluorogenic Peptide Cleavage Assay to Screen for Proteolytic Activity: Applications for coronavirus spike protein activation. JoVE, e58892.

Kawaoka, Y., and Webster, R.G. (1988). Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. Proceedings of the National Academy of Sciences *85*, 324-328.

Le Coupanec, A., Desforges, M., Meessen-Pinard, M., Dube, M., Day, R., Seidah, N.G., and Talbot, P.J. (2015). Cleavage of a Neuroinvasive Human Respiratory Virus Spike Glycoprotein by Proprotein Convertases Modulates Neurovirulence and Virus Spread within the Central Nervous System. PLoS Pathog *11*, e1005261.

Lee, D.W., and Whittaker, G.R. (2017). Use of AAScatterPlot tool for monitoring the evolution of the hemagglutinin cleavage site in H9 avian influenza viruses. Bioinformatics *33*, 2431-2435.

Licitra, B.N., Millet, J.K., Regan, A.D., Hamilton, B.S., Rinaldi, V.D., Duhamel, G.E., and Whittaker, G.R. (2013). Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. Emerg Infect Dis *19*, 1066-1073.

Millet, J.K., Goldstein, M.E., Labitt, R.N., Hsu, H.-L., Daniel, S., and Whittaker, G.R. (2016). A camel-derived MERS-CoV with a variant spike protein cleavage site and distinct fusion activation properties. Emerging Microbes & Infections *5*, e126.

Millet, J.K., and Whittaker, G.R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proceedings of the National Academy of Sciences *111*, 15214–15219.

Nao, N., Yamagishi, J., Miyamoto, H., Igarashi, M., Manzoor, R., Ohnuma, A., Tsuda, Y., Furuyama, W., Shigeno, A., Kajihara, M.*, et al.* (2017). Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin. MBio *8*.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J*., et al.* (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun *11*, 1620.

Shapiro, J., Sciaky, N., Lee, J., Bosshart, H., Angeletti, R.H., and Bonifacino, J.S. (1997). Localization of Endogenous Furin in Cultured Cell Lines. Journal of Histochemistry & Cytochemistry *45*, 3-12.

Steinhauer, D.A. (1999). Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology *258*, 1-20.

Straus, M.R., and Whittaker, G.R. (2017). A peptide-based approach to evaluate the adaptability of influenza A virus to humans based on its hemagglutinin proteolytic cleavage site. PLoS One *12*, e0174827.

Tse, L.V., Hamilton, A.M., Friling, T., and Whittaker, G.R. (2014). A Novel Activation Mechanism of Avian Influenza Virus H9N2 by Furin. Journal of Virology *88*, 1673-1683.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell.

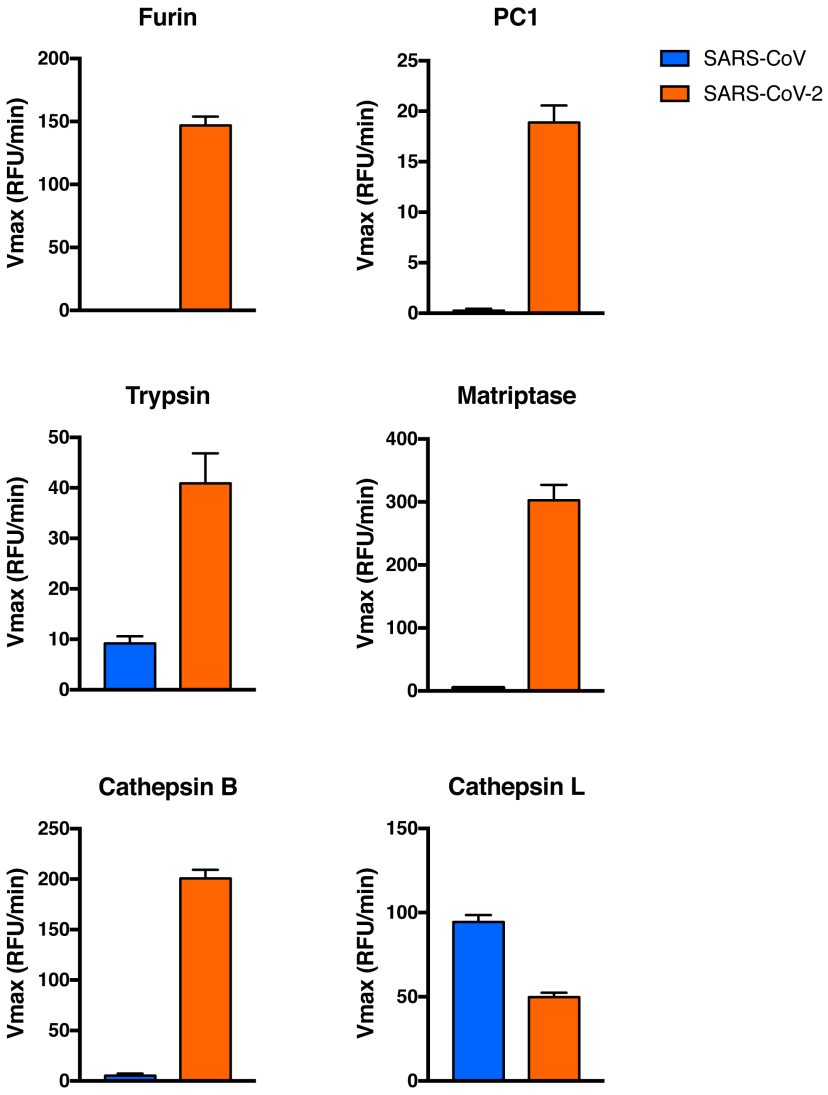
Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C*., et al.* (2020a). A novel bat coronavirus reveals natural insertions at the S1/S2 cleavage site of the Spike protein and a possible recombinant origin of HCoV-19. bioRxiv, 2020.2003.2002.974139.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L*., et al.* (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270-273.

248 **Figure 1**

249



250